ONTARIO

MULTIVARIATE ANALYSIS AND RESIDENTIAL PROPERTY VALUATION IN ONTARIO

Methodology Section
Assessment Standards Branch
Ontario Department of Municipal Affairs

October 1970

THE HONOURABLE W. DARCY McKEOUGH
MINISTER

ONTARIO DEPARTMENT OF MUNICIPAL AFFAIRS

W. H. PALMER
DEPUTY MINISTER

ONTARIO DEPARTMENT OF MUNICIPAL AFFAIRS

P. G. GILLIS
EXECUTIVE DIRECTOR
ASSESSMENT DIVISION

ONTARIO DEPARTMENT OF MUNICIPAL AFFAIRS

MULTIVARIATE ANALYSIS AND RESIDENTIAL

PROPERTY VALUATION IN ONTARIO

Methodology Section
Assessment Standards Branch
Ontario Department of Municipal Affairs

This paper describes research in the statistical valuation of residential property undertaken by the Ontario Government.* The research was begun in early 1970, and until recently it has been concerned largely with problems of data collection and methodology. Consequently the paper tends to focus on basic conceptual issues and preliminary empirical analysis, rather than on conclusive "results".

The paper is divided into four parts. Part one briefly describes the governmental setting of our activity. Part two outlines a conceptual approach to the use of multivariate analysis in the valuation of residential property. Part three illustrates some preliminary applications of this approach using data on single-family and multi-family dwellings. To conclude, part four offers some suggestions about both procedural issues and the likely direction of our future research.

PART ONE: GOVERNMENTAL SETTING

In January 1970 the Government of Ontario -- analogous to state government in the United States -- assumed responsibility for the assessment of the local property tax base, as part of a general program for the reform of municipal finance. Before this date real property assessment in the province had been administered through the system of local government.

Since the Second War property assessment in Ontario has fallen into a critical state. Basic shortcomings have been magnified by the rapid growth and shift of population, together with the growing cost and complexity of municipal government.

The need for assessment reform was stressed in the Report of the Ontario Committee on Taxation (1967), which noted: "There has been a growing tendency that extreme inadequacies in property assessment with resulting inadequacies in taxation, have been hidden from view by the prevalence of gross under assessment". For example, in 1966 assessment levels were below 40% of current values in 822 of the 935 municipalities in the province.

This Report prompted the Ontario Government to institute a system of centralized property assessment at current "market value", administered by the provincial civil service. Naturally the striking of tax rates and the collection of taxes remains with local governments. The chief reasons for this centralization involved a general quest for equity, and the need to improve the

base for the provincial-local transfer payments which have become critical in the last two decades.

Since January 1970 property assessment in Ontario has been administered by the Assessment Division of the Department of Municipal Affairs. The Division is responsible for the valuation of approximately 3,750,000 properties, and current policy calls for a complete reassessment by 1975. For purposes of management the province is divided into thirty-two operating regions and seven administrative areas. An Assessment Standards Branch and Assessment Education Branch in Toronto provide technical advice and assistance.

The work on statistical valuation presented in this paper has been developed by the Methodology Section of the Standards Branch, in company with resource persons from other parts of the Ontario Government. This work is oriented to the development of computer-based statistical valuation as a practical tool in property tax assessment.

PART TWO: A CONCEPTUAL APPROACH

SCOPE AND GENERAL APPROACH

The scope of our current work involves an attempt to apply some statistical techniques associated with multivariate analysis to the valuation of single-family and multi-family dwellings in urban centres. For our purposes an urban centre is an industrially or commercially based community of at least 50,000 people. (See map page 5) By limiting the scope of our initial research we do not imply that multivariate analysis is unsuited to the valuation of other classes of property, or of property in other types of community. We have simply chosen urban dwellings as the focus of our current work.

Within this framework our objective is to predict the "market value" of individual properties with a degree of accuracy appropriate to property-tax assessment. At this point in our work just what this degree of accuracy should be is undecided. As a rough and ready guide we regard predictions within $\pm$ 10% of "observed value" with a reliability level of 95% as suitable for research and development.

For us "market value" means "most likely sale price". Our reason for choosing this standard of value is the principle of equity in tax assessment. Roughly, a particular property can be treated as one in a group of physically similar properties. Ordinarily when properties within this group sell they do not all bring the same price. Instead they sell within a range of prices, and in a general sense most likely sale price is the average of this range.

POPULATIONS OVER 50,000
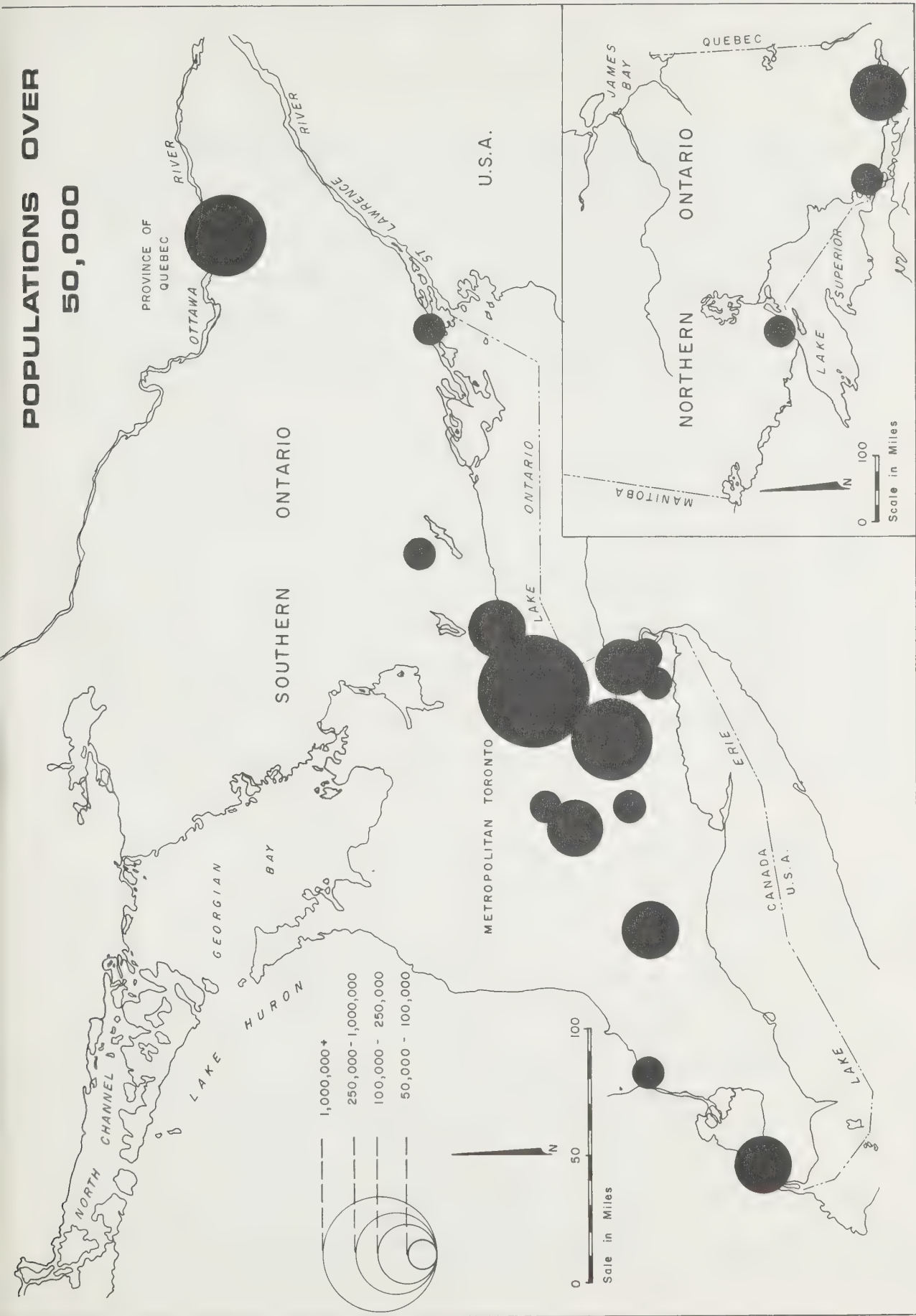
We regard value in this sense as a function of consumer or investor behaviour. We assume that this function is reflected in the relations between sale price and various physical and locational characteristics of recently sold properties. By establishing these relations through statistical analysis, we hope to predict the value of similar properties that have not sold.

Given this general framework, it seems clear that there are basic problems associated with any attempt to use statistical techniques in the valuation of real property. These problems relate to differences between the material that statistical tecnhiques are designed to work with, and the material that real property markets make available for analysis. Although at this point in our work we are not clear on how these problems can be most easily resolved, it seems useful to outline briefly what we think they are.

Broadly speaking statistical methods allow us to make inferences about relations in a "population" on the basis of how the same relations behave in a "sample". Naturally inferences of this sort are only possible when the "units" in the sample adequately represent the units in the population with regard to the relations under study. For example, if we wished to study the relation between personal income and education we would first define our population, and then select a sample from this population so that each income level and educational background was represented by enough observations to ensure reliable inferences.

This type of sampling is possible because we can collect information on both income and education for any unit in our population.

Unfortunately this procedure cannot be rigorously applied in statistical approaches to real property valuation. For example in the case of single-family dwellings our "ideal" population naturally includes all single-family dwellings in a particular city. Yet we can only collect information about sale price for those properties which have sold. Consequently we cannot choose a sample so that all single-family dwellings in a city are adequately represented with regard to those characteristics that have an important influence on sale price. Our best sample is all single-family dwellings that have sold in the recent past, and this sample is "forced" on us by real estate markets. In this context the most we can do is ensure that we make inferences only about those kinds of properties that are adequately represented by recent sales.

This raises two important points. First, it seems clear that because of scanty sales data, there will be some types of single-family or multi-family dwelling that cannot be valued using the techniques associated with multivariate analysis. This weakness, of course, is common to any valuation technique that uses market information. Second, we are interested in representativeness not with regard to any characteristic, but only with regard to those characteristics that are statistically significant for predicting sale price. Because of this we cannot specify exactly what our population is until we have completed our analysis of recent sales.

This point might be clarified by viewing statistical valuation as an attempt to apply the traditional "market approach" to mass appraisal conditions. In this context the potential of multivariate analysis is that it seems to provide a systematic framework for making the subjective decisions associated with the adjustment of "comparables" on a scale suited to mass valuations. Yet just as the market approach cannot be used when few good comparables for a particular property are available, statistical valuation is limited by the extent of recent sales data. Our work so far suggests that large numbers of the urban dwellings in a typical centre are amenable to some form of statistical valuation. Nonetheless distinguishing between those that are and those that are not remains a critical task.

Viewing statistical valuation as a mass-appraisal adaptation of the traditional market approach has an additional implication for our work that is closely related to the problem of unrepresented properties. Just as we cannot value small bungalows by comparing them with sales of large three-storey houses, it seems clear that we will require separate multivariate models for different groups of relatively similar single-family or multi-family dwellings (Smith, 1970, p 8). In this context an important part of our research involves stratifying our sample of sales into component "markets", "groups", or "market aggregations", in a way that will allow us to develop models which accurately predict the value of individual properties.

We noted earlier that we are studying the relations between sale price and the physical and locational characteristics of recently sold properties. Our reasons for examining only these characteristics are twofold. First, physical and locational characteristics are readily available in most assessment offices, and second they seem most appropriate in the context of valuations for the real property tax.

By proceeding in this way we assume that characteristics which have an influence on value but are not directly associated with actual properties (or with the details of their exchange) -- e.g. characteristics of occupants, like income or occupation -- are adequately reflected in the relations between sale price and physical and locational characteristics. For example, we assume that the socio-economic characteristics which distinguish between values in different neighbourhoods are reflected by broad location variables. As our analysis proceeds we might find that this type of assumption is unwarranted. In general, however, we hope to confine our models to characteristics directly associated with real property.

STATISTICAL METHODS

Property valuation using multivariate techniques has been studied in other areas with varying degrees of success (e.g. Gustafson, 1967 and Shenkel , 1968). The most common procedure has been to employ multiple regression analysis with sale price as the dependent variable and a selection of physical property

characteristics as independent variables. There are four major
problems associated with this approach. First, the set of
variables chosen by the researcher to describe the population must
be sufficiently comprehensive to include the major factors
affecting market value. Secondly, the regression model may contain
independent variables which are highly related, causing collinearity
effects. Third, the form of the model is generally postulated as a
simple linear additive function which does not adequately reflect
interactions that may be inherent in the variables. Fourth, the
data may violate conditions of normality and linearity implicit
in the usual multiple regression models.

We feel that even with its inherent weaknesses multiple
regression analysis is the best means to estimate the final models
for the population parameters lying within the range of our sample
data. Our approach, however, will attempt to minimize these
deficiencies by using several methods in the initial analysis of
the data.

We propose to use histograms and simple graphical
representations combined with Chi-square analysis to discover
non-linear relationships within the data, and transformations which
might be necessary in order to correct for deviations from
normality. Secondly, we will use factor analysis to identify the
underlying logical structure of our data in order to isolate data
gaps, and obtain a better understanding of the interdependencies
among the variables used to measure property characteristics. In
addition, we will employ "Automatic Interaction Detection" (AID) to try
to discover interaction effects between variables and to provide a

means of stratifying the data into homogeneous groups for use in subsequent regression sub-models.

Finally, a step-wise multiple regression procedure will be employed to reduce multi-collinearity effects. In each regression model, qualitative characteristics will be dichotomized into (0, 1) variables in order to eliminate problems arising from arbitrarily assigning scales to the intervals which are not measurable in absolute units. Residuals will be examined to discover non-linear trends and to identify any property types in the sample which have anomalous characteristics that might require further study.

In summary, it is our objective to use simple tabulations and histograms, AID and factor analysis in a preliminary exploration to evaluate the nature of the data, determine its scope of application, and identify areas of multicollinearity and interaction. Then on the basis of this knowledge, we intend to develop multiple regression models which are structured to maximize predictive relationships, by minimizing both the inclusion of spurious terms and the exclusion of significant factors, and by combining the factors so as to represent the interactions identified in the preliminary analysis.

PART THREE: EMPIRICAL ILLUSTRATIONS, SINGLE-

FAMILY AND MULTI-FAMILY DWELLINGS

For our first major study of both single-family and
multi-family dwellings we selected the City of Oshawa, a city of
roughly 85,000 people slightly to the east of Metropolitan
Toronto (see map p.13).

Most of our attention in this study has been directed to
single-family dwellings. In this context the material that
follows presents some illustrations of the statistical methods
discussed above. These illustrations relate to preliminary
analysis rather than the formulation of final models.

Naturally the number of multi-family sales in a city of
Oshawa's size is limited. This has meant that preliminary
multivariate analysis of these sales was not possible since both
factor analysis and AID require a fairly large number of observations.

Nonetheless we felt that it was useful to examine the limited
multi-family data available. In the first place we gained exposure
to problems associated with the collection and validation of
multi-family data. Second, working with multi-family sales gave us
an opportunity to become familiar with our stepwise regression
program, while conducting preliminary analysis of single-family sales.

SINGLE FAMILY

Data Base

In Oshawa all sales of single-family dwellings during 1969 were
collected by field staff in the Ontario County Assessment Region.
After "non-valid" transactions had been rejected we were left with
a sample of 1553 valid sales. Sales data was then collated with
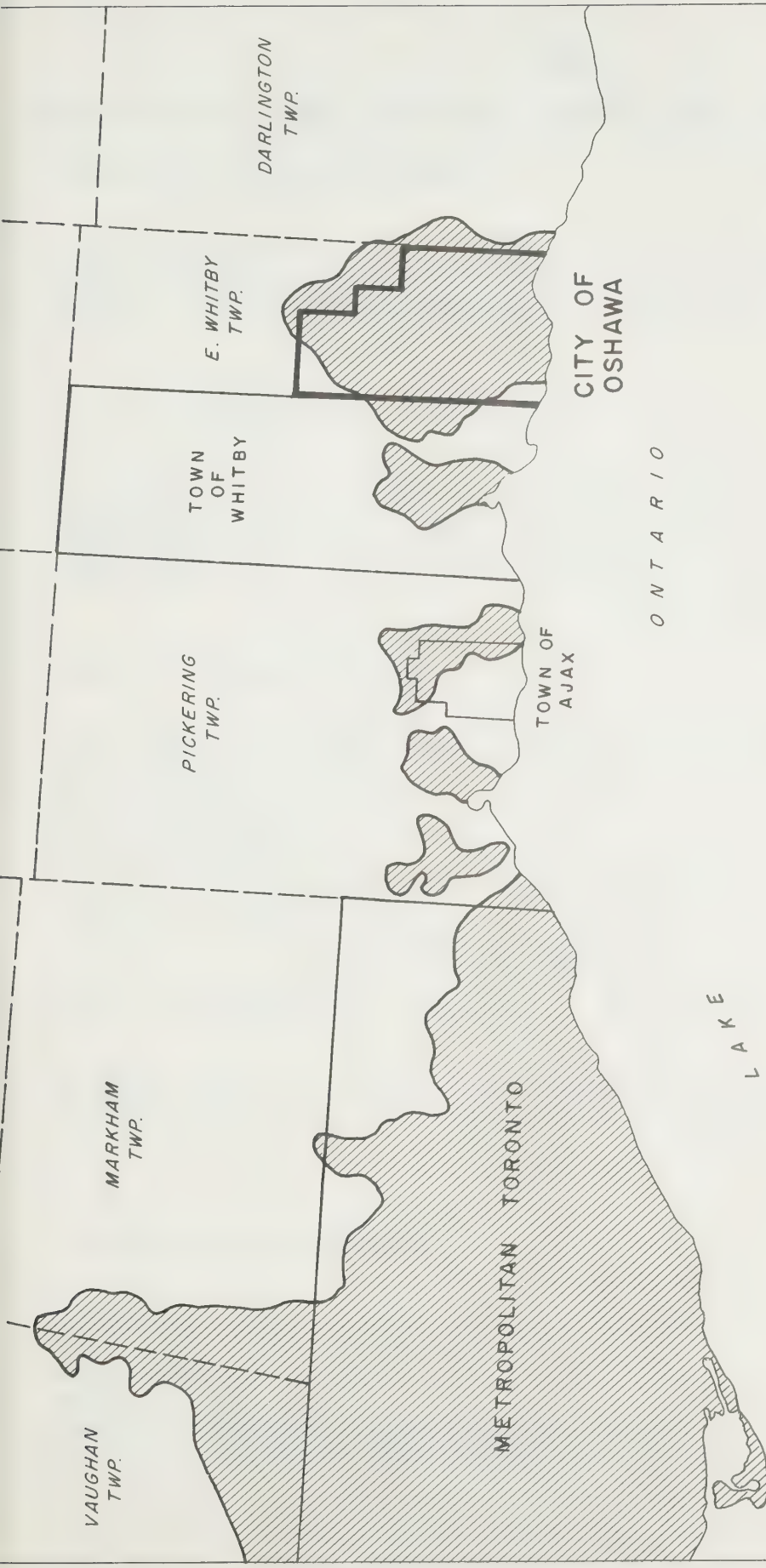
LOCATION OF STUDY AREA

——————— Study Area

Approximate Developed Urban Area

DARLINGTON TWP.

E. WHITBY TWP.

CITY OF OSHAWA

TOWN OF WHITBY

PICKERING TWP.

TOWN OF AJAX

MARKHAM TWP.

VAUGHAN TWP.

METROPOLITAN TORONTO

L A K E   O N T A R I O

N

0   5

Scale in Miles
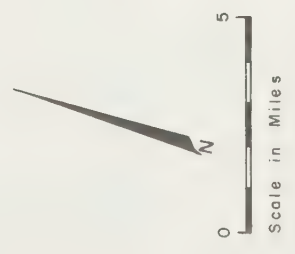
TABLE 1

SINGLE-FAMILY CHARACTERISTICS, SALES DATA, CITY OF OSHAWA 1969

TRANSACTION CHARACTERISTICS

Total Sale Price.....................................Dollars
Date of Sale...Months
Ratio of Cash to Total Price ....Decimal fraction

SITE CHARACTERISTICS

Site Area ....Sq. Feet
Site Frontage..Feet
Site Depth...Feet
Site Shape...Regular or
                                                    Irregular

BUILDING CHARACTERISTICS

Total Floor Area.....................................Sq. Feet
No. of Rooms
No. of Bedrooms
No. of Bathrooms
No. of Stories
Building Design......................................Detached or
                                                    Semi-Detached
Building Structure...................................Brick, Frame,
                                                    Brick Veneer,
                                                    Other Concrete
Building Quality.....................................Substandard to
                                                    Excellent (nine
                                                    classes)
Building Age...Years
Garage...None, Carport,
                                                    Single, Double

LOCATION CHARACTERISTICS

Community..Nine Areas, see
                                                    map, p. 15
Neighbourhood
Block
Distance Downtown....................................Feet
Distance 401 (Major Throughway)......................Feet

OSHAWA
COMMUNITY
BOUNDARIES

1

2

SIMCOE ST.

3

4

5

C.B.D.

6

7

8

KING ST.

HWY. 401

9

LAKE ONTARIO

N

0 Scale in Miles 1
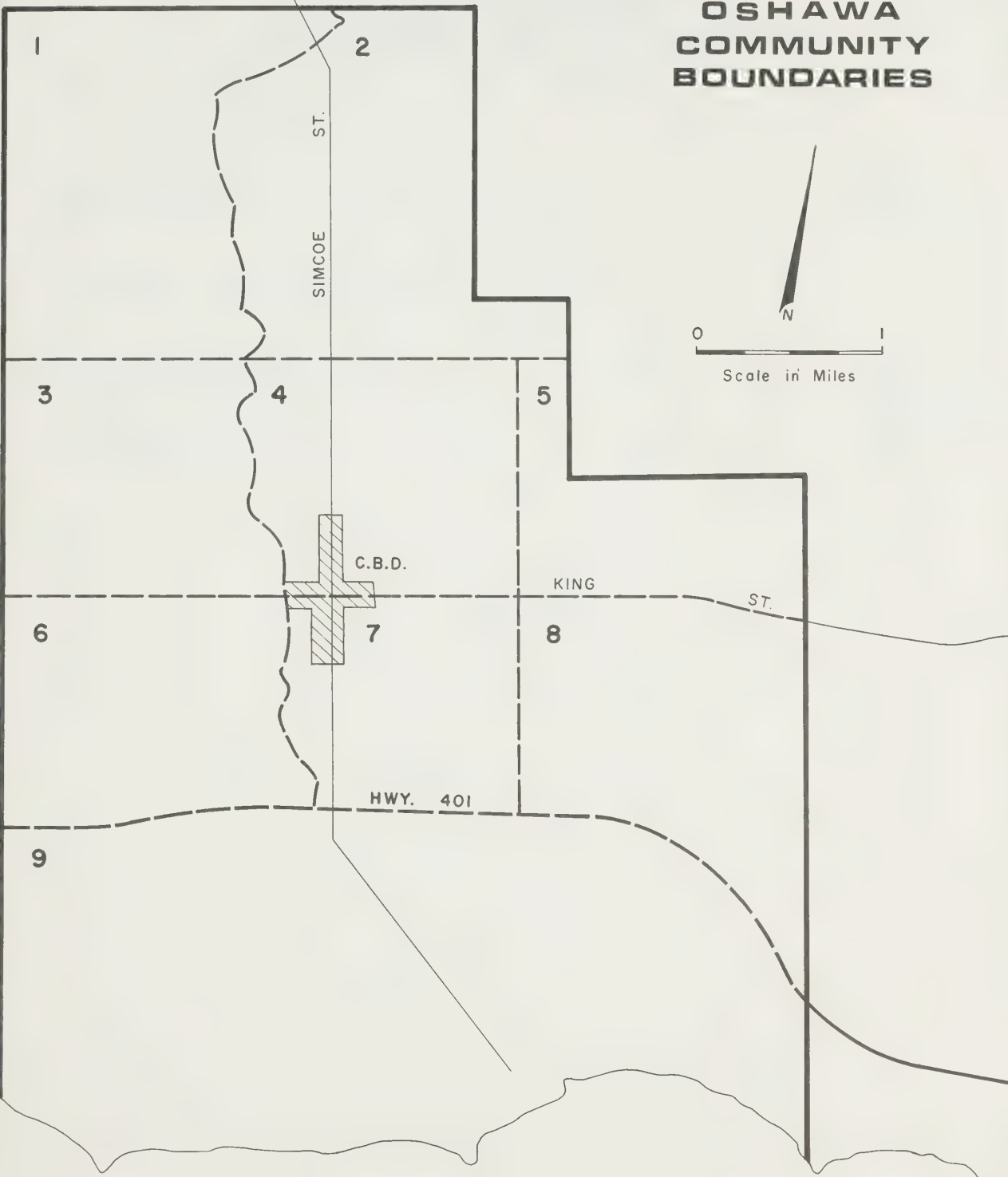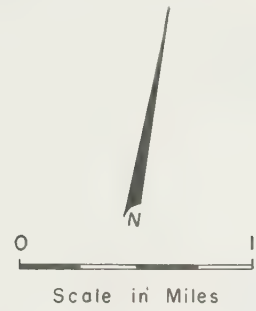
TABLE 2
SINGLE-FAMILY CHARACTERISTICS
ALL PROPERTIES, CITY OF OSHAWA 1969.

SITE CHARACTERISTICS

Site Frontage..Feet
Site Depth..Feet

BUILDING CHARACTERISTICS

Total Floor Area.......................................Sq. Feet
Building Design..Detached or
                                                       Semi-detached
Building Age...Years

appraisal cards to determine physical and locational characteristics.
The set of characteristics that we collected is outlined in
Table 1 (p. 13).

In addition to sales or sample data we also collected some
information on all single-family dwellings in the City of Oshawa,
to provide possible assistance in specifying the scope of application
for our predictive models. The characteristics collected in this
case are set out in Table 2 (p. 16).

Graphical Analysis

As we noted earlier, we are using simple graphical techniques
to explore problems associated with deviations from normality. In
this context, as part of our early analysis we developed
histograms for each of the characteristics set out in Tables 1 and 2.
In some cases histograms for sample characteristics exhibited
considerable skewness. For example, the histogram for sale price
was skewed somewhat to the right, and plotting the sale-price
frequency distribution on probability paper indicated that a
logarithmic transformation might be appropriate for analysis of the
total sample. On this basis we included a transformation of sale
price in our early regression runs on the unstratified sample of
Oshawa sales. Naturally this transformation might not be
appropriate within strata of relatively similar properties.

To look more closely at problems related to normality, we have
recently acquired a Chi-Square program which calculates the
parameters of several theoretical distributions and compares them
with the distributions of observed data. The statistics generated
by this program suggest types of transformations that might correct

for deviations from normality. Although we have not yet used this program, we hope that it will improve the reliability of our final predictive models.

Factor Analysis

So far we have performed one factor analysis on our complete Oshawa sales sample using the 31 variables set out in Table 3. These variables were generated from the sample characteristics of Table 1. As indicated in Table 3 several characteristics were dichotomized in (0, 1) variables.

This analysis produced 13 orthogonal factors which were subsequently rotated using a varimax rotation. Briefly, the 13 factors are totally uncorrelated, and those variables with high loadings on any one factor are, to a high degree, "substitutes" for that factor. Floor area, rooms, and bedrooms, for example, were highly loaded on the first factor, with floor area having the greatest communality. This suggests that these three variables all express a single factor readily identified as "building size". It also suggests that of the three variables floor area is likely to be the most successful index of building size.

In several cases community variables had high loadings on other significant factors which could not be readily identified. This might simply reflect different socio-economic conditions in various parts of the city. On the other hand, it is also possible that there are property characteristics closely associated with community but not explicitly measured in our present data. In this context one possibility suggested by regional assessment staff is the "traffic obsolescence" created by the proximity of major transportation arteries.

TABLE 3

FORM OF ENTRY OF SINGLE-FAMILY DATA IN

FACTOR ANALYSIS AND MULTIPLE REGRESSION ANALYSIS,

CITY OF OSHAWA, 1969

TRANSACTION CHARACTERISTICS

Total Sale Price.....................................Dollars
Date of Sale..Months

SITE CHARACTERISTICS

Site Frontage.......................................Feet
Site Depth..Feet
Site Shape..Regular or
                                                    Irregular

BUILDING CHARACTERISTICS

Total Floor Area....................................Sq. Feet
No. of Rooms
No. of Bedrooms
No. of Bathrooms
No. of stories
Building Design.....................................Detached or
                                                    Semi-detached
Building Structure
    Brick...0, 1
    Other Concrete..................................0, 1
    Brick Veneer....................................0, 1
    Frame...0, 1
Building Age..Years
Garage
    No garage or carport............................0, 1
    Carport...0, 1
    Single Garage...................................0, 1
    Double Garage...................................0, 1

LOCATION CHARACTERISTICS

    Community
        1...0, 1
        2...0, 1
        3...0, 1
        4...0, 1
        5...0, 1

LOCATION CHARACTERISTICS (Cont'd)

    Community
        6..0, 1
        7..0, 1
        8..0, 1
        9..0, 1
Distance Downtown...Feet
Distance 401 (Major Throughway)................................Feet

Automatic Interaction Detection

"Automatic Interaction Detection" (AID) is a statistical technique for computers developed at the University of Michigan's Institute for Social Research. A basic description of the program is set out in The Detection of Interaction Effects (Sonquist and Morgan 1964). More detailed treatment appears in Multivariate Model Building (Sonquist, 1970). We will not attempt to explain the technique in detail, since this is beyond the scope of the paper. For our purposes here, AID splits sample data into groups of relatively similar properties on the basis of differences in mean sale prices.

So far we have run our Oshawa sample data through the AID program twice. In our first run we used the eighteen characteristics set out in Table 4. On the basis both of this information and of insights gained from our factor analysis we performed a second run using only eleven characteristics. In this second run we excluded rooms and bedrooms since our factor analysis indicated that they index the same type of characteristic as floor area. Similarly, we excluded characteristics like site shape and distance from downtown because our initial AID run suggested that they were not closely related to sale price. The eleven characteristics used in our second run are set out in Table 5.

The "tree" which appears on p. 26 presents the results of our second AID run. Each box in the tree represents a number of sales that share the characteristics noted in the box, along with any characteristics in preceding boxes on the same "branch" of the tree. The first number in a box is the mean price of the sales (in

TABLE 4

SAMPLE CHARACTERISTICS USED IN FIRST AID RUN

CITY OF OSHAWA, SINGLE FAMILY.

Sale Price	No. of bathrooms
Date of Sale	No. of stories
	Building Design
Site Frontage	Building Structure
Site Depth	Building Quality
Site Shape	Building Age
	Garage
Total Floor Area	
No. of rooms	Community
No. of bedrooms	Distance Downtown

Distance 401

TABLE 5

SAMPLE CHARACTERISTICS USED IN SECOND AID RUN,

CITY OF OSHAWA, SINGLE FAMILY.

Sale Price	No. of Bathrooms
Date of Sale	Building Design
	Building Age
Site Frontage	Garage
Site Depth	
	Community
Total Floor Area	Distance 401

thousands of dollars), the second is the standard deviation of
the prices (again in thousands of dollars), and the third is the
actual number of sales. For example, box 45 - in the middle part
of the page - represents sales which share the characteristics of
building age 40 years or more, one bathroom only, community 3-9
(see map. p. 15) and floor area more than 1,400 sq. ft. The mean
price of these sales is $21,700, the standard deviation is $4,800
and there are 26 sales in the box. Roughly speaking, the tree as
a whole is set out so that the highest mean prices are at the top
and the lowest at the bottom, while those in between follow more or
less in sequence.

As we noted earlier, we hope to use AID both to help detect
interactions and to suggest meaningful strata for the development
of predictive models. In the tree set out on p. 26 interaction
might be illustrated by the branches located in the bottom, left-hand
corner of the page. In boxes 8-9, 16-19, for example, it appears
that the relation between sale price and floor area varies with
building age. Comparison of boxes 17 and 18 indicates that the
mean price of smaller but newer houses is almost the same as the
mean price of larger but older houses. This suggests that a
variable which combines floor area and building age might result
in more accurate predictions than we could achieve by using the two
characteristics independently. Some suggestions about the
construction of this type of variable are set out in Multivariate
Model Building, p.p. 209-213

Turning to our second use of the AID technique, it is possible that the construction of "interaction terms" might create almost as many problems as it solves. In this context attempting to avoid interactions through stratification might prove more useful. Naturally this approach is also attractive since it seems clear that stratification of our sample data is desirable for the more general reasons noted earlier (see p. 8). As we see it now, different branches of the AID tree might serve as preliminary strata for testing regression models. The branch represented by box 25, for example, suggests a stratum ("market", "group", or "market aggregation") of semi-detached houses with no garage or carport, less than 14 years old, and floor areas of 1400 square feet or less.

The AID tree as a whole might also be used to give some indication about specifying those parts of the single-family population that can be valued using multivariate techniques. Clearly sales of the largest properties are in short supply. For example, we have only 24 sales of two bathroom houses in Communities 1 and 2 with floor areas over 1400 sq. ft. This is not enough observations to develop a reliable regression model. Although we are not clear about what implications this might have for the scope of our final models, it seems quite possible that we will have trouble with large houses. In this context data collected in Table 2 indicates that over 85% of all single-family dwellings in Oshawa have floor areas of 1400 sq. ft. or less. In other words, even if we cannot value any of the properties represented by those boxes in the top half of the AID tree, our

models might still apply to over 85% of the single family population.

So far as we know, AID has had little previous application in real estate analysis. In fact even in other fields the technique is recent enough to make reliable judgements about its behaviour difficult. In our own work we are not very familiar with the interpretation of AID or with its real potential and limitations. Given this background of uncertainty, the extent to which some of the splits on p.26 "make sense" is encouraging. The data divides initially, for instance, on floor areas more or less than 1400 sq. ft. This seems appropriate since in southern Ontario 1400 sq. ft. is also the typical minimum building size in those areas where local governments have attempted to encourage "estate development". Similarly, the splits on building age more or less than 40 years that appear in boxes 16-19 and boxes 44 and 45 conform with the scatter diagram of sale price and building age set out on p. 27 This would seem to support a common-sense distinction between houses built before and after the Great Depression.

Regression Analysis

We have recently run regressions on our sample of all single-family sales using the 31 variables set out in Table 3. As we suggested earlier we used a logarithmic transformation of sale price in these regressions, to correct for deviations from normality. As one might expect these regression runs suggest that general regressions on all sales are not suitable as predictive models. Nonetheless we feel that detailed analysis of the runs might suggest important non-linear trends or help to identify property types which require further study.

SECOND A.I.D. RUN

SINGLE FAMILY DWELLINGS

CITY OF

OSHAWA

RELATIONSHIP OF AGE OF DWELLING TO SALES PRICE

SINGLE FAMILY DWELLINGS, OSHAWA, 1969
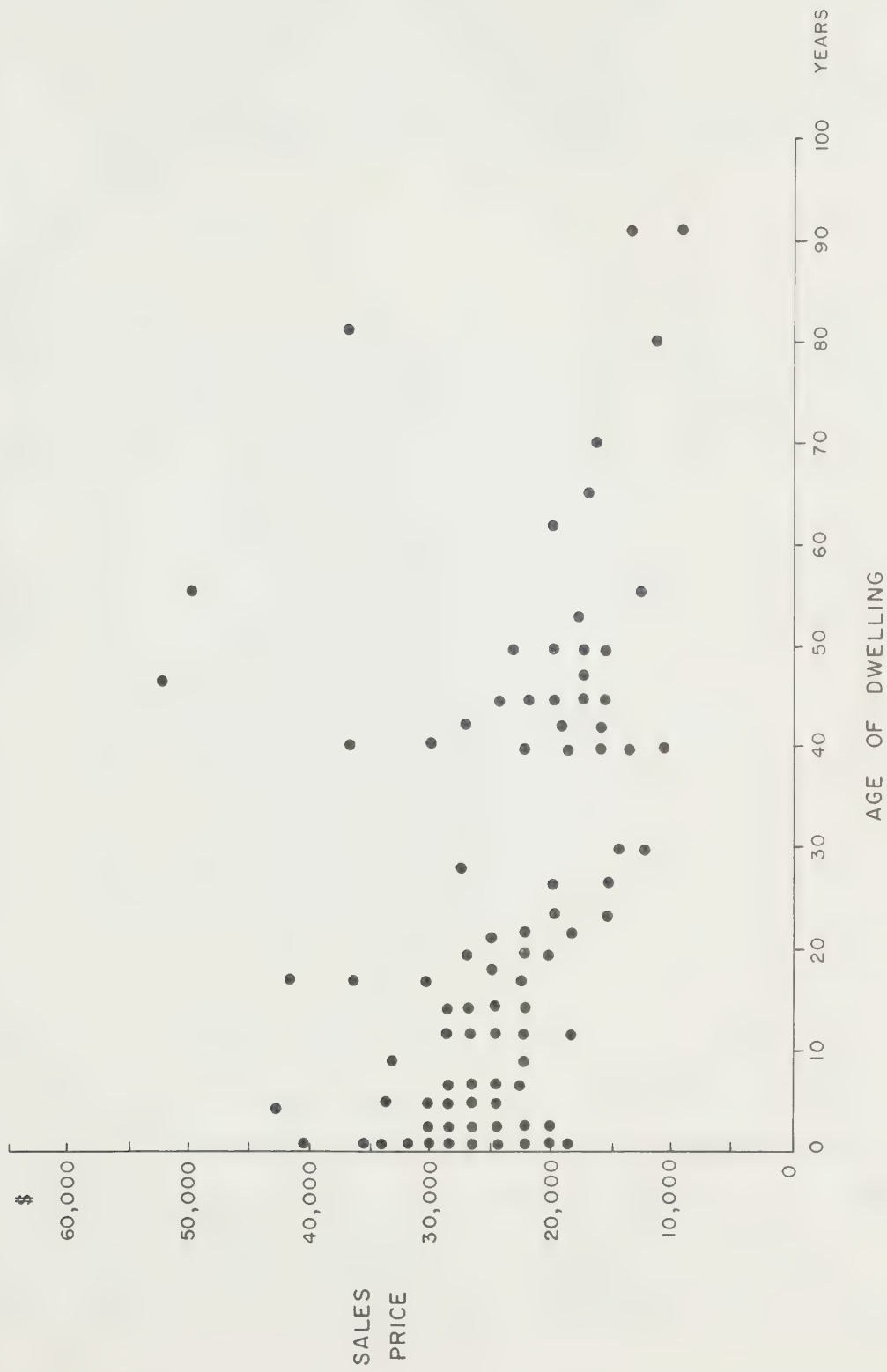
Naturally our next step is to run separate regressions on different strata of relatively similar properties within our total sample. The results of our general regressions are encouraging enough to suggest that this step should produce some acceptable predictive models.

MULTI FAMILY

Data Base

Our multi-family study began with a small sample of 72 sales which included properties ranging in size from 2 to 108 suites. This represents all sales during the two-year period 1968-1969. We selected this length of time on the assumption that two years is a stable market period in a city of Oshawa's size. Forty-nine of the properties were "apartments", and twenty-three were "conversions", originally designed for single-family use.

Transaction characteristics were obtained from the local registry office and merged with property characteristics from appraisal cards in the Regional Assessment Office. The characteristics that we collected are set out in Table 6 (p.29).

As noted earlier one reason for studying multi-family data was to develop a sales validation procedure for use in later projects. This involved a careful examination of sale price, gross income per suite, and gross income multipliers. On this basis properties which appeared to involve unusual transactions were inspected in the field and checked at both registry and assessment offices. As a result of this procedure eight apartments and one conversion were eliminated from analysis, leaving a working sample of 63 properties.

TABLE 6

MULTI-FAMILY CHARACTERISTICS

SALES DATA, CITY OF OSHAWA, 1968-69

TRANSACTION CHARACTERISTICS

No. of Mortgages
Total Sale Price..Dollars
Date of Sale...Months
Grantor..Individual or
                                                       Corporation
Grantee..Individual or
                                                       Corporation

RENTAL CHARACTERISTICS

Gross Annual Rent......................................Dollars
Effective date of rent.................................Months
Part owner occupied or fully rented

BUILDING CHARACTERISTICS

Type...Conversion or
                                                       Apartment
Exterior Finish:
    Brick 0, 1
    Brick Veneer.......................................0, 1
    Reinforced Concrete................................0, 1
Age..Year of
                                                       Construction
No. of Storeys
No. of bathrooms
No. of units
Type of Heating:
    Forced Air...0, 1
    Hot Air..0, 1
    Stove..0, 1
    Hot Water..0, 1
    Steam..0, 1
    Electricity..0, 1
Ground Floor area of building..........................Sq. feet
Gross floor area.......................................Sq. feet

TABLE 6 (Cont'd)

SITE CHARACTERISTICS

Frontage..Feet
Depth..Feet
Area...Sq. Feet
Shape..Regular or
                                                      Irregular

LOCATION CHARACTERISTICS

    Community
          1...0, 1
          2...0, 1
          3...0, 1
          4...0, 1
          5...0, 1
          6...0, 1
          7...0, 1
          8...0, 1
          9...0, 1
    Distance Downtown ....Feet
    Distance 401 (Major Throughway).........................Feet

Regression Analysis

We first applied step-wise multiple regression to all 63 valid sales of multi-family dwellings, and then performed a second analysis of the apartments alone. In both cases we used the variables set out in Table 6. Similarly, in both cases we used a logarithmic transformation of sale price to correct for deviations from normality.

Since preliminary multivariate analysis could not be performed on our small amount of multi-family data, it is difficult to make inferences about any regression results. This difficulty is increased by the large amount of variation in our multi-family sample. Nonetheless we feel that the interpretations we have attempted are worth noting, since they appear to have some implications for the development of predictive models in areas where more data is available.

Our regression analysis of all 63 sales suggests that several of the characteristics we collected account for a large amount of variation in the prices of multi-family dwellings in our data. The coefficient of determination, for example, is 97.5% and the "relative" standard error of estimate* is 3.8%.

The residuals from regression, expressed as a percentage of observed values, were randomly distributed with respect to the dependent variable. Ignoring sign, the mean of the residuals was 13% and the highest value was 42%.

The step-wise procedure treated the following variables as significantly related to sale price:

---

* Standard error of estimate ÷ mean sample sale price.

Variable	Student "t" of Regression Coefficient $(t_{.90}, df = 50)$
Type (conversion or apt.)	6.9
No bathrooms	6.3
Ground floor area	5.9
Total floor area	5.3
Brick finish	3.2
Part owner occupied	2.9
No mortgages	2.7
Community 7	2.6
Date of Sale	2.2
Shape of lot	2.2
Community 6	1.9
Grantee	1.7

Regression analysis of the 41 apartment sales alone produced almost identical regression statistics. The coefficient of determination, for example, was the same as in the first run, while the relative standard error of estimate dropped slightly to 3.2%. Again, residuals from regression were randomly distributed, with a mean of 13%. In this case, however, the highest value was only 34%.

The step-wise procedure treated the following variables as significantly related to sale price:

Variable	Student "t" of Regression Coefficient $(t_{.90}, df = 32)$
No bathrooms	6.5
Ground floor area	5.3
Total floor area	5.0
Brick finish	2.7
Forced air heating	2.3
Date of sale	2.3
No mortgages	2.3
Shape of lot	1.8

As one might expect, our first regression suggests that characteristics distinguishing between conversions and apartments account for a considerable amount of the variation in price within our complete sample of multi-family sales. In addition to building type, other variables appear to be making this distinction. For example, more conversions than apartments are part owner occupied, and most conversions are located in Communities 7 and 8, older parts of the city near the central business district.

This interpretation seems to be confirmed by our analysis of apartments alone. In this case, apart from building type -- which naturally did not appear in our second run -- the three variables noted above were treated as insignificant by the step-wise procedure.

Both regressions suggest that number of mortgages and date of sale are significantly related to sale price. These variables present some problems for predictive models, since they cannot be determined for unsold properties. In this context several approaches seem possible. Even in a city as small as Oshawa, for instance (at least when it is located in a rapidly growing area), it seems that the stable market period for multi-family dwellings is less than two years. On this basis it would seem wise to collect sales from a somewhat shorter period, or to try building some type of adjustment factor into predictive models.

The influence of financing on sale price, of course, is a traditional appraisal problem. The results of our regression analysis suggest that in this area it might be useful to employ accepted "discounting" techniques, or again to experiment with adjustment factors in predictive models.

PART FOUR: PROCEDURAL PROBLEMS AND RESEARCH PROSPECTS

Throughout our research we have had difficulty with both data collection and computer facilities. Our data collection problems have resulted from the need to convert information designed for other purposes into a form suited to multivariate analysis. These problems seem unavoidable in the context of the administrative arrangements which prevail in most assessment offices. Information on replacement cost, for example, is not ideal for market analysis using statistical methods. Similarly, in many cases, sales data is not integrated with information on property characteristics, and few offices have any information suited to immediate data processing for analytic purposes.

As part of our work we are interested in the development of a complete support system for statistical valuation. At the same time it seems clear that work on multivariate analysis for mass appraisals has not progressed to a point where it is possible to specify a completely suitable data system. In this context the best locations for multivariate research are likely those jurisdictions which have well-organized property records and accurate sales data, regardless of form. This consideration played a large part in our selection of Oshawa as an initial study area. For the time being, however, any multivariate analysis is likely to be coloured by the nature of available data, which in most cases is designed for slightly different purposes.

In broad perspective, problems associated with computer facilities have created more delays for us than those related to data collection. It seems that the basic reason for these problems

has been our lack of direct access to the computer. Having access only through an outside programming staff has meant, for example, that we cannot perform even simple data manipulations without significant delays created by programming time. It has also meant that our "turnaround" time is often longer than we would like.

There seem to be two approaches to these problems. On the one hand, we are attempting to link some related programs to minimize the number of separate computer runs. For example, we are attempting to link simple sorting and tabulation routines with our basic multivariate programs. On the other hand, we have had some experience with terminal units, and it seems likely that many of our problems would be resolved by this type of facility.

The key to successful valuation using multivariate techniques is the ability to make inferences from sold to unsold properties. It seems clear that this requires an intensive preliminary analysis designed to provide well-defined predictive models whose scope of application can be clearly identified. In this context our future research will attempt to extend the type of preliminary empirical work that appears in this paper.

Naturally we will also begin the formulation of predictive models where our existing analysis suggests that this might be appropriate. At the moment, however, we feel that we are some distance away from the type of final model that could confidently be used in a practical valuation system.

Once we feel that we have exhausted the potential of our present data from the City of Oshawa, we intend to begin new

studies in other parts of the province. Hopefully this will give
us some idea of problems associated with the use of
multivariate techniques in different geographical areas. It
should also help suggest which parts of our work in Oshawa apply to
residential property markets in general and which parts apply to
Oshawa alone.

	# Suite	# Rooms	Average Suite Size	Price	Income	G.I.M.	Average Rent per suite P.M.	Price per suite	Remarks
Conversions									
1	2	5	2.5	23180	2700	8.5	112	11590	
2	2			25000	2640	9.5	100	12500	
3	3	6	2.0	23000	3600	6.4	100	7667	
4	3	8	2.6	26000	3480	7.5	97	8666	
5	3	7	2.3	24000	3600	6.6	100	8000	
6	3	7	2.3	25000	3180	7.8	88	8333	
7	3	8	2.6	30500	3600	8.5	100	10166	
8	3	6	2.0	26000	3420	7.6	95	8666	
9	3	8	2.6	28500	4620	6.2	128	9500	
10	3	7	2.3	25500	3756	6.7	104	8500	
11	3	7	2.3	52000	3420	15.2	96	17333	
12	3	6	2.0	27200	-	-	-	9070	
13	3	7	2.3	23000	-	-	-	7666	
14	4	12	3.0	33000	5760	5.7	120	8250	
15	4	8	2.0	45000	5040	8.9	105	11250	
16	4	9	2.2	23800	4200	5.7	87	5950	
17	4	8	2.0	42500	4740	8.9	98	10625	
18	4	12	3.0	22500	4500	5.0	94	5625	
19	4	8	2.0	19000	3276	5.8	67	4750	
20	5	10	2.0	53000	3360	15.7	56	10600	
21	5	11	2.0	53000	-	-	-	10600	
22	8	14	1.7	33000	8604	3.8	89	4125	Exclude
23	11	27	2.4	107500	14604	7.4	86	9772	
Apartments									
24	2	7	2.6	87500				43750	"
25	3	8	2.6	29500	4320	6.8	120	9833	"
26	3	8	2.6	29500	4440	6.4	123	9833	"
27	3	8	2.6	38500	4560	8.4	126	12833	
28	3	8	2.6	35500	3780	9.4	105	11833	
29	3	11	3.6	33500	5460	6.1	151	11166	
30	5	15	3.0	54000	6636	8.1	110	10800	

# Suite	# Rooms	Average Suite Size	Price	Income	G.I.M.	Average Rent per suite P.M	Price per suite	Remarks
31	5	3.0	25600	6780	3.7	113	5000	Exclude
32	5	3.0	54900	6420	8.6	107	10981	
33	5	3.0	45000	6840	6.6	114	9000	
34	6	2.8	58000	15396	3.8	214	9666	⌠ Correct
35	6	2.6	81000	10920	7.4	151	13500	⌡ Income
36	6	2.8	70250	9600	7.3	133	11708	
37	6	2.8	71000	10860	6.5	150	11833	
38	6	3.6	65000	9480	6.9	131	10833	
39	6	2.8	67500	8220	8.2	114	11250	
40	6	2.6	61000	7176	8.5	99	10166	Include
41	6	2.6	60500	7656	7.9	106	10083	
42	6	2.6	63000	7740	8.1	107	10500	
43	6	2.5	38000	7500	5.1	104	6333	Exclude
44	6	2.8	64500	9840	6.6	136	10750	
45	8	1.5	34000	7000	4.6	72	4250	Correct inc.
46	8	2.3	32100	11900	2.7	124	4012	Exclude
47	8	2.2	111000	11604	9.6	120	13875	
48	9	2.8	93750	16478	5.7	145	10418	Include
49	9	2.8	91500	13752	6.7	127	10166	
50	9	2.7	88000	13020	6.8	120	9777	
51	10	3.0	120000	15060	8.0	125	12000	
52	11	3.0	105000	15346	6.8	116	9549	
53	11	2.4	105000	16020	6.6	121	9545	
54	11	3.0	115000	15432	7.5	116	10434	
55	11	2.4	107500	15060	7.1	114	9772	Exclude
56	11	2.4	110000	15000	7.1	113	10000	
57	11	2.7	126000	16200	6.1	122	11454	
58	11	3.0	116905	10740	7.0	126	10627	
59	11	3.0	116905	15840	7.4	120	10627	
60	12	2.0	250000	10800	23.1	75	20833	Exclude
61	15	2.2	135000	16200	8.3	90	9000	
62	18	3.0	140000	20560	6.8	95	7777	
63	20	2.5	267000	33456	8.0	139	13350	
64	20	2.5	244000	34920	7.0	145	12200	
65	23	2.5	259000	39288	6.6	142	11260	

# Suite	# Rooms	Average Suite Size	Price	Income	G.I.M.	Average Rent per suite P.M	Price per suite	Remarks
66	92	2.8	305000	53448	5.7	139	9531	
67	272	3.1	875000	153168	5.7	148	10174	
68	280	3.0	1049000	161016	6.5	145	11402	Include
69	282	3.0	1070000	157824	6.8	142	13178	
70	245	2.5	1252000	162744	7.7	142	13178	Include
71	322	3.0	1162565	179160	6.5	139	10865	Include
72	334	3.0	1340000	201444	6.7	155	12407	Include

REFERENCES

Gustafson, Robert H., "E.S.P. and the Appraiser", Address delivered
    at the 35th Annual Conference of the National Association of
    Tax Administrators, San Francisco, 14 June 1967.

Shenkel, Dr. William M., "Computer Valuation By Multiple Regression
    Analysis", Proceedings of the 34th Annual International Conference
    on Assessment Administration (Chicago, 1968), 26-39.

Smith, T. Reynolds, "Linear Regression Analysis of Sales Data",
    Assessors Journal, July 1970, 3-9

Sonquist, John A. and Morgan, James N., The Detection of Interaction
    Effects (Ann Arbor, 1964).

Sonquist, John A., Multivariate Model Building (Ann Arbor, 1970).